PENERAPAN ALGORITMA BIDIRECTIONAL ENCODER FROM TRANSFORMER UNTUK MENDETEKSI PLAGIARISME PADA KODE PROGRAM BAHASA C++

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika

disusun oleh

BAYU TAUFIQURRAHMAN

18.11.2086

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

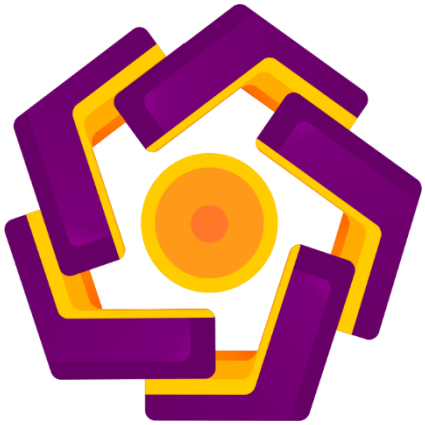
2023

PENERAPAN ALGORITMA BIDIRECTIONAL ENCODER FROM TRANSFORMER UNTUK MENDETEKSI PLAGIARISME PADA KODE PROGRAM BAHASA C++

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Informatika



disusun oleh

BAYU TUFIQURAHMAN

18.11.2086

Kepada

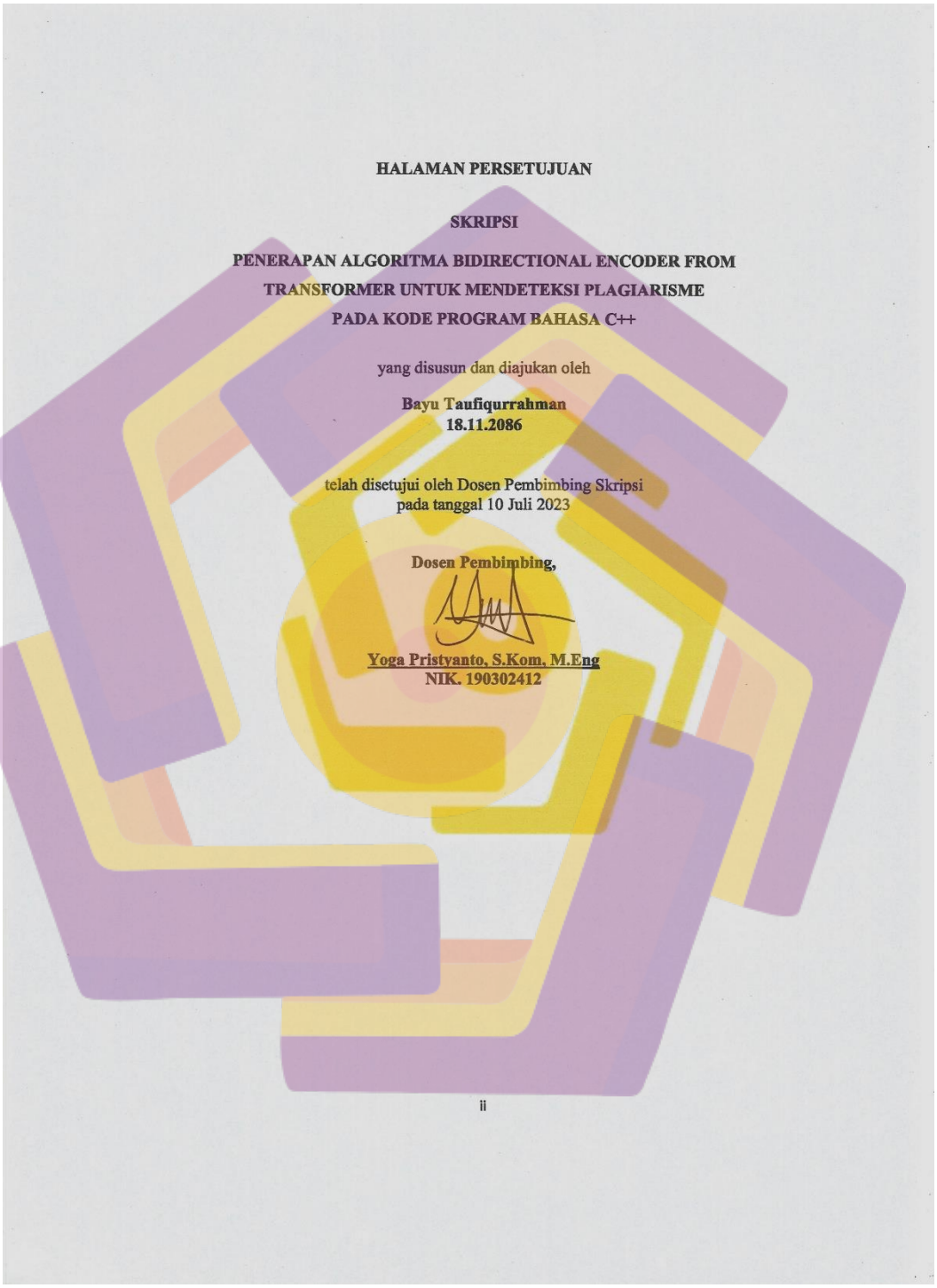
FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2023

HALAMAN PERSETUJUAN



HALAMAN PERSETUJUAN

SKRIPSI

PENERAPAN ALGORITMA BIDIRECTIONAL ENCODER FROM TRANSFORMER UNTUK MENDETEKSI PLAGIARISME PADA KODE PROGRAM BAHASA C++

yang disusun dan diajukan oleh

Bayu Taufiqurrahman
18.11.2086

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 10 Juli 2023

Dosen Pembimbing,

Yoga Pristyanto, S.Kom, M.Eng
NIK. 190302412

ii

HALAMAN PENGESAHAN

HALAMAN PENGESAHAN

SKRIPSI

PENERAPAN ALGORITMA BIDIRECTIONAL ENCODER FROM TRANSFORMER UNTUK MENDETEKSI PLAGIARISME PADA KODE PROGRAM BAHASA C++

yang disusun dan diajukan oleh

Bayu Taufiqurrahman
18.11.2086

Telah dipertahankan di depan Dewan Penguji
pada tanggal 31 Juli 2023

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Yoga Pristyanto, S.Kom, M.Eng **NIK. 190302412**	
Dr. Ferry Wahyu Wibowo, S.Si., M.Cs. **NIK. 190302235**	
Majid Rahardi, S.Kom., M.Eng **NIK. 190302393**	

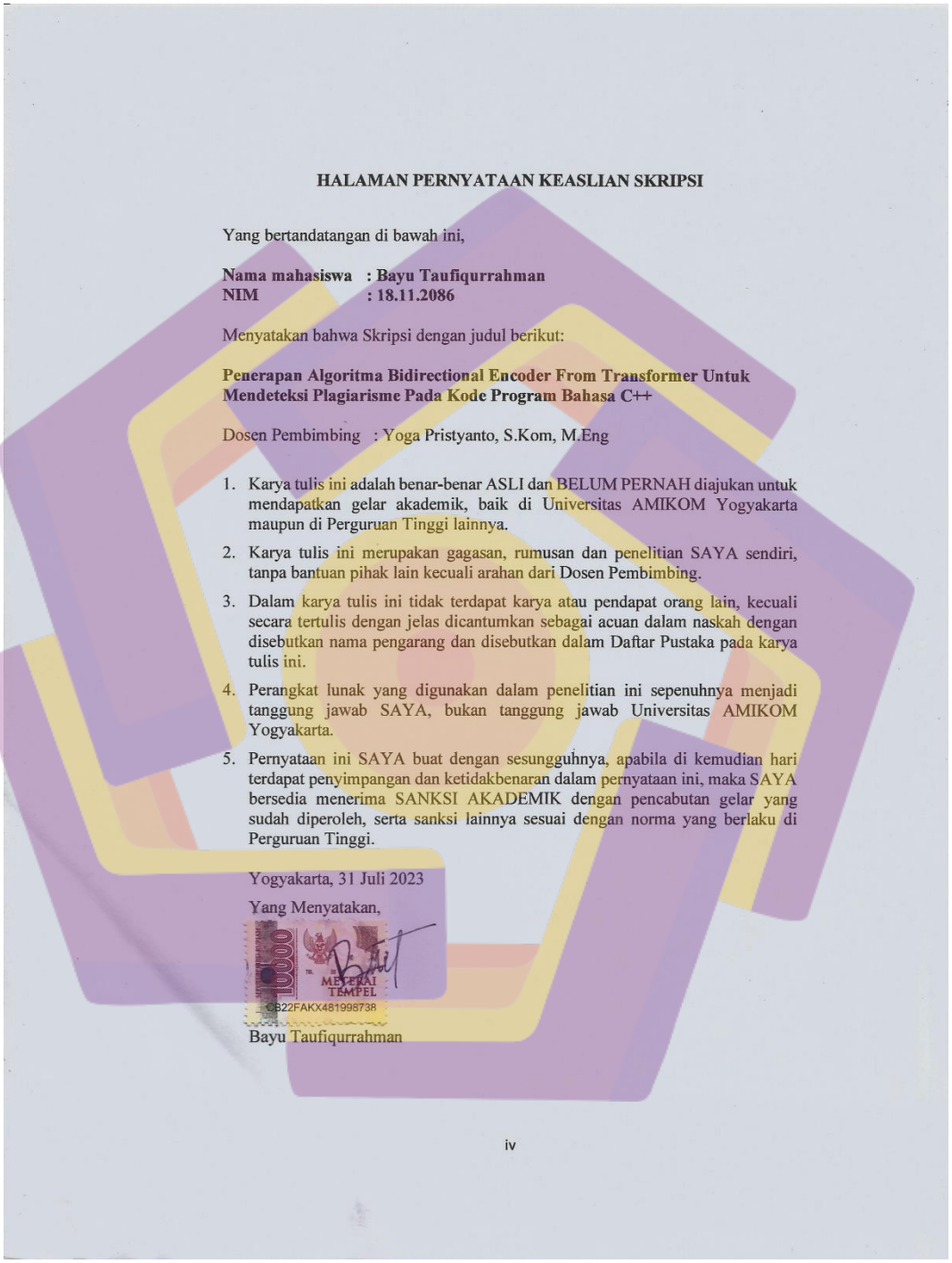
Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 31 Juli 2023

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta,S.Kom., M.Kom.
NIK. 190302096

iii

HALAMAN PERNYATAAN KEASLIAN SKRIPSI



HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Bayu Taufiqurrahman
NIM : 18.11.2086

Menyatakan bahwa Skripsi dengan judul berikut:

Penerapan Algoritma Bidirectional Encoder From Transformer Untuk Mendeteksi Plagiarisme Pada Kode Program Bahasa C++

Dosen Pembimbing : Yoga Pristyanto, S.Kom, M.Eng

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 31 Juli 2023

Yang Menyatakan,

Bayu Taufiqurrahman

iv

KATA PENGANTAR

Assalamu'alaikum warahmatullahi wabarakatuh.

Puji Syukur kehadirat Allah SWT atas limpahan rahmat dan hidayah-Nya, penulis dapat menyelesaikan tugas akhir ini dengan judul "Penerapan Algoritma Bidirectional Encoder From Transformer Untuk Mendeteksi Plagiarisme Pada Kode Program Bahasa C++". tugas akhir ini merupakan salah satu persyaratan yang harus peneliti penuhi untuk memperoleh gelar sarjana Fakultas Ilmu Komputer Jurusan Informatika di Universitas Amikom Yogyakarta. Selama proses penulisan tugas akhir ini, penulis menghadapi berbagai hambatan, namun berkat dukungan, bimbingan, dan kerjasama yang tulus dari berbagai individu, akhirnya tugas akhir ini berhasil diselesaikan dengan sukses. penulis mengucapkan terima kasih kepada:

1. Bapak Yoga Pristyanto, S.Kom, M.Eng selaku dosen pembimbing yang telah memberikan bimbingan, arahan, motivasi serta kesabaran kepada penulis selama pengerjaan tugas akhir sehingga tugas akhir ini dapat terselesaikan.
2. Bapak Dr. Ferry Wahyu Wibowo, S.Si., M.Cs., Bapak Majid Rahardi, S.Kom., M.Eng selaku dosen penguji yang telah memberikan saran untuk perbaikan tugas akhir ini.
3. Kedua orang tua serta wali penulis yang selalu memberikan doa, motivasi dan dukungan moral dalam menyelesaikan tugas akhir ini.
4. Seluruh jajaran dosen dan staff Universitas Amikom Yogyakarta.
5. Teman-teman IF04 dan rekan kerja student staff DAAK Pengajaran yang saling mendukung.

Penulis memahami bahwa penelitian dalam tugas akhir ini masih jauh dari kesemurnaan, oleh karena itu, penulis mengundang pembaca untuk memberikan kritik dan saran. Harapan penulis adalah agar hasil dari tugas akhir ini dapat memberikan nilai dan manfaat yang signifikan bagi semua pihak yang terlibat.

Yogyakarta, 31 Juli 2023

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

INTISARI

Plagiarisme merupakan sebuah tindakan yang mengambil karya, ide, ataupun informasi tanpa sepengetahuan pemilik hak cipta kemudian mengakuinya atas miliknya sendiri. Tugas akhir pemrograman setiap semester rentan terhadap plagiarisme oleh mahasiswa. Untuk mendeteksi adanya perbuatan plagiarisme, penelitian ini menggunakan algoritma bidirectional encoder from transformer. Algoritma bidirectional encoder from transformer diperkenalkan oleh Google untuk meningkatkan ketepatan mesin pencarian miliknya.

Penelitian ini dilakukan melalui beberapat tahap yaitu (1) data cleansing tujuannya untuk membuang data yang tidak dapat terpakai agar tidak menyebabkan overfitting maupun underfitting. (2) Visualisasi data merupakan tahap mengetahui data yang digunakan untuk training. (3) Training dan testing data merupakan sebuah proses untuk melatih dan mengetahui hasil dari algoritma yang diterapkan.

Dataset yang digunakan untuk penelitian ini berasal dari IBM dengan total seribu empat ratus kode program bahasa C++ yang setiap kodenya sudah dilabeli dengan waktu proses cpu, dan penggunaan memori yang digunakan saat program dijalankan. Dataset ini akan dilakukan preprocessing terlebih dahulu seperti data cleansing dan tokenizing. Hasil dari penelitian diharapkan dapat membuat para pengajar lebih mudah untuk mengatasi plagiarisme kode program bahasa C++.

Kata Kunci: bert, nlp, transformer, machine learning, plagiarisme

ABSTRACT

Plagiarism is an act of taking works, ideas, or information without the knowledge of the copyright owner and then acknowledging it as his own. Final project on programming per semester is easily plagiarized by students. To detect plagiarism, this research uses a bidirectional encoder from transformer algorithm. The bidirectional encoder from transformer algorithm was introduced by Google to improve the accuracy of its search engine.

This research was carried out through several stages, namely (1) data cleansing, which was to dispose of data that could not be used so as not to cause overfitting or underfitting. (2) Data visualization is the stage of knowing the data used for training. (3) Training and data testing is a process to train and find out the results of the applied algorithm.

The dataset used for this research comes from IBM with a total of one thousand four hundred C++ language program codes, each of which has been labeled with cpu processing time, and the memory usage used when the program is run. This dataset will be preprocessed first, such as data cleansing and tokenizing. The results of the research are expected to make it easier for teachers to overcome plagiarism in the C++ language programming.

Keyword: bert, nlp, transformer, machine learning, plagiarism